

IOWA STATE UNIVERSITY

Digital Repository

Industrial and Manufacturing Systems Engineering
Conference Proceedings and Posters

Industrial and Manufacturing Systems Engineering

2006

Optimal Solution to a Capacity Expansion Problem

Rahul R. Marathe
Iowa State University

Sarah M. Ryan
Iowa State University, smryan@iastate.edu

Follow this and additional works at: https://lib.dr.iastate.edu/imse_conf



Part of the [Operations Research, Systems Engineering and Industrial Engineering Commons](#)

Recommended Citation

Marathe, Rahul R. and Ryan, Sarah M., "Optimal Solution to a Capacity Expansion Problem" (2006). *Industrial and Manufacturing Systems Engineering Conference Proceedings and Posters*. 135.
https://lib.dr.iastate.edu/imse_conf/135

This Conference Proceeding is brought to you for free and open access by the Industrial and Manufacturing Systems Engineering at Iowa State University Digital Repository. It has been accepted for inclusion in Industrial and Manufacturing Systems Engineering Conference Proceedings and Posters by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.

Optimal Solution to a Capacity Expansion Problem

Abstract

For a service provider, stochastic demand growth along with expansion lead times and economies of scale may encourage delaying the start of expansion until after some shortages have been accumulated. Assuming demand follows a geometric Brownian motion, we define the service level in terms of the proportion of demand satisfied, which is then analytically evaluated using financial option pricing theory. Under a stationary expansion policy, an infinite time horizon discounted expansion cost is minimized under the service level constraint, where the expansion timing and size parameters are the decision variables. With the current formulation, the problem seems to be unbounded.

Keywords

Capacity expansion, service level, barrier option pricing, cutting plane algorithm

Disciplines

Operations Research, Systems Engineering and Industrial Engineering

Comments

This is a proceeding published as Rahul R. Marathe and Sarah M. Ryan; Optimal Solution to a Capacity Expansion Problem. IIE Annual Conference. Proceedings 2006. Posted with permission.

Optimal Solution to a Capacity Expansion Problem

Rahul R. Marathe and Sarah M. Ryan; Department of Industrial & Manufacturing Systems Engineering; Iowa State University; Ames, IA 50011-2164, USA.

Abstract

For a service provider, stochastic demand growth along with expansion lead times and economies of scale may encourage delaying the start of expansion until after some shortages have been accumulated. Assuming demand follows a geometric Brownian motion, we define the service level in terms of the proportion of demand satisfied, which is then analytically evaluated using financial option pricing theory. Under a stationary expansion policy, an infinite time horizon discounted expansion cost is minimized under the service level constraint, where the expansion timing and size parameters are the decision variables. With the current formulation, the problem seems to be unbounded.

Keywords

Capacity expansion, service level, barrier option pricing, cutting plane algorithm

1. Introduction

Capacity expansion problems arise in numerous applications varying from communications networks to manufacturing facilities. The problem is to find an optimal policy of expansion given a particular forecasted demand pattern, assuming that the costs and lead times of expansion are known.

We consider a service provider having certain facilities with installed capacity to provide certain services. We consider a single location and single resource assuming that the demand for that resource follows a geometric Brownian motion (GBM) process. The capacity added does not deteriorate; that is, once the capacity is installed, we assume that it is available forever. Expansion costs exhibit economies of scale and there is a deterministic expansion lead time from the time the capacity expansion decision is made to the time when the added capacity is actually available to satisfy the demand.

Modeling demand as a GBM process may be justified when empirical data show that demand growth in a period is on average a constant percentage of demand at the beginning of the period, and periods of higher or lower than average demand occur at random. Marathe and Ryan [1] verified empirically that the historical usage of electric power in the US as well as the number of passenger enplanements in the airline industry each followed a GBM process.

The capacity expansion literature is richly stocked. Manne [2] considered a random-walk pattern demand and proposed the optimal size of the capacity expansion when there were economies of scale available. Whitt [3] considered the utilization aspect of the capacity expansion and found the stationary distribution for the capacity utilization under a simple policy that we adapt in this paper. Chaouch and Buzacott [4] considered the demand with plateaus and formulated the capacity expansion for two cases, viz., when the expansion starts with some initial shortages and when it starts before the demand reaches the current capacity. Our model is similar to this case, with our demand being GBM process driven. Bean et al. [5] considered demand to be following either a transformed Brownian motion process or a semi-Markovian birth and death process. They showed that the problem can be transformed into an equivalent deterministic problem and that the effect of the probabilistic nature of demand is to reduce the interest rate. This result was used by Ryan [6] wherein the effect of a fixed lead time was also considered. Financial option pricing theory was used to develop a stationary expansion policy so that the specified service level is met when the expansion started before the demand reaches the current capacity position. Our model is further extension of Ryan [6] in the sense that we consider the case where the expansion starts when the demand has already crossed the current capacity position.

Our parameter definitions are similar to Ryan [6]; and most of the details may be found in Marathe and Ryan [7, 8]. We briefly summarize the model and definitions in section 2; the service level and the expansion cost analyses are carried out in section 3. We present a numerical example in section 4 and conclude the paper with section 5.

2. Model

Demand for the service is given by the GBM process $P(t) = P(0)e^{B(t)}$, where $B(t)$ is a Brownian motion with drift μ and variance σ^2 . Define $\gamma \equiv \mu + \frac{\sigma^2}{2}$ as the mean (exponential) growth rate of the demand.

We assume that capacity additions occur at discrete time points and that a fixed lead time of L time units is required to install new capacity. The problem is to choose a sequence $\{T_n, X_n\}_{n \geq 1}$, where T_n , the time when the n^{th} capacity expansion starts, is a stopping time with respect to the Brownian motion $B(t)$ and X_n is the n^{th} increase in capacity. For any realization ω of the Brownian motion $B(t)$, let $t_n \equiv T_n(\omega)$. Let K_n be the installed capacity after n additions are completed, where the initial capacity is K_0 . The capacity position includes capacity on order (being constructed or installed) in addition to the installed capacity.

We describe the model by quoting directly from Marathe and Ryan [7, 8]. We follow the Whitt-Luss policy from Whitt [3], where each new expansion starts when demand reaches some fixed proportion (say, ' p ') of current capacity position, and after its addition at the end of the lead-time, the new capacity is a constant proportion of its previous value. That is $K_n = \nu K_{n-1}$, where $\nu > 1$. For the case of $p < 1$ Ryan [6] used financial option pricing theory to find optimal stationary expansion policy (that is, the values of parameters p and ν). In this paper, we consider the case where $p \geq 1$.

Figure 1 illustrates the policy and potential shortages seen at the realized time t_n , when demand first equals pK_{n-1} . The n^{th} capacity expansion has just started. With this expansion, the total installed capacity will reach level K_n after the lead time L . As stated earlier, we model the situation wherein the service provider waits until certain amounts of initial capacity shortages are accumulated before starting the next expansion project. Hence, since the new capacity position is K_n , the next expansion would start at the time when the demand $P(t)$ first reaches the position pK_n . Since the demand process is stochastic, this time for starting the next expansion (T_{n+1}) is a random variable. The goal, then, is to find the optimal initial shortage that will trigger the start of capacity expansion (the parameter p), and the optimal size of each expansion (the parameter ν). Figure 1 shows a non-overlapping expansion cycle where the capacity being built is already available before we begin the next expansion. It is also possible for expansion cycles to overlap.

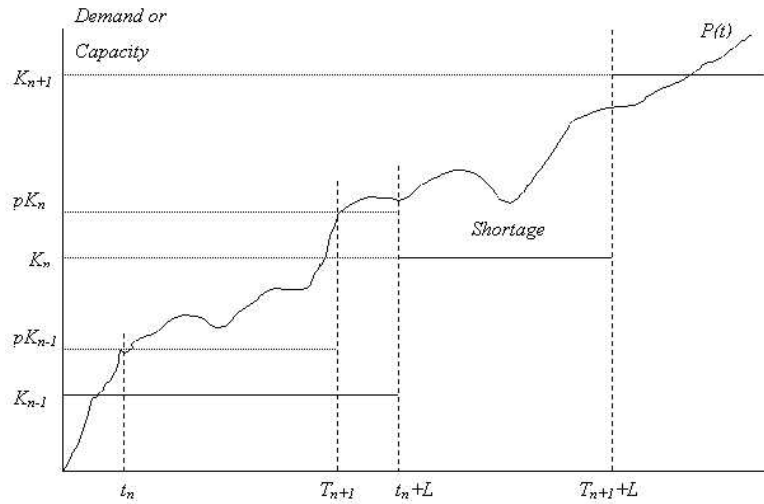


Figure 1. Capacity expansion policy when the expansion starts after the end of the current expansion cycle.

As explained in Marathe and Ryan [7], the service level in the expansion cycle $[t_n + L, T_{n+1}]$ is defined as:

$$\beta = E \left[\frac{\int_{t_n}^{T_{n+1}} m \left[\frac{P(\cdot)}{P(\cdot)} \right] t \, dK}{\int_{t_n}^{T_{n+1}} P(\cdot) t \, dK} \right], \quad d \left[\frac{P(\cdot)}{P(\cdot)} \right] t \, dK \quad (1)$$

After approximations, the above equation becomes:

$$\beta(p, v) = 1 - \frac{E \left[\int_{t_n}^{T_{n+1}} P(\cdot) t \, dK \right]}{p \left[E_{n+1} T_n \right]}. \quad (2)$$

We note that the service level is the same for each expansion cycle and a function only of the policy parameters p and v .

3. Analysis of Service Level and Expansion Cost

By comparing the numerator of Equation (2) with the up-and-out barrier option and also simplifying the denominator, we have the capacity shortage equation as:

$$\begin{aligned} 1 - \beta(p, v) &= \frac{I}{p \left[E_{n+1} T_n \right]} \quad \text{where} \\ I &= \int_L^\infty e^{(\gamma-r)u} \left\{ \left(\frac{p}{v} \right) \Psi \left(\frac{-1 \left(\frac{v}{p} \right) \gamma \left(\frac{\sigma^2}{2} \right) u - 1 \left(\frac{v}{p} \right) \gamma \left(\frac{\sigma^2}{2} \right) L}{\sigma \sqrt{u}}, \frac{1 \left(\frac{v}{p} \right) \gamma \left(\frac{\sigma^2}{2} \right) L}{L \sqrt{u}} \right) \sqrt{\frac{u-L}{u}} \right. \\ &\quad - v^{\frac{2\gamma}{\sigma^2}+1} \left(\frac{p}{v} \right) \Psi \left(\frac{-1 \left(\frac{v}{p} \right) \gamma \left(\frac{\sigma^2}{2} \right) u - 1 \left(\frac{v}{p} \right) \gamma \left(\frac{\sigma^2}{2} \right) L}{\sigma \sqrt{u}}, \frac{1 \left(\frac{v}{p} \right) \gamma \left(\frac{\sigma^2}{2} \right) L}{L \sqrt{u}} \right) \sqrt{\frac{u-L}{u}} \\ &\quad \left. - e^{-\gamma u} \Psi \left(\frac{-1 \left(\frac{v}{p} \right) \gamma \left(\frac{\sigma^2}{2} \right) u - 1 \left(\frac{v}{p} \right) \gamma \left(\frac{\sigma^2}{2} \right) L}{\sigma \sqrt{u}}, \frac{1 \left(\frac{v}{p} \right) \gamma \left(\frac{\sigma^2}{2} \right) L}{L \sqrt{u}} \right) \sqrt{\frac{u-L}{u}} \right. \\ &\quad \left. + e^{-\gamma u} \frac{2\gamma}{\sigma^2} \left(\frac{p}{v} \right) \Psi \left(\frac{-1 \left(\frac{v}{p} \right) \gamma \left(\frac{\sigma^2}{2} \right) u - 1 \left(\frac{v}{p} \right) \gamma \left(\frac{\sigma^2}{2} \right) L}{\sigma \sqrt{u}}, \frac{1 \left(\frac{v}{p} \right) \gamma \left(\frac{\sigma^2}{2} \right) L}{L \sqrt{u}} \right) \sqrt{\frac{u-L}{u}} \right\} \quad (3) \end{aligned}$$

and $\Psi(x, y)$ is the bivariate normal distribution function evaluated at (x, y) with correlation coefficient ρ .

To evaluate the infinite time horizon total cost of expansion, let $V_t(K)$ be the minimum expected cost, at time t with capacity position K , of expanding capacity over infinite horizon while satisfying the service level constraint. Let the rate at which future costs are discounted be r . Referring to Figure 1, at time t_n , when the expansion has just been initiated, our goal is to find the timing (p) and size (v) parameters for the next expansion. We assume an economies of scale regime, under which cost of installing capacity of size X is given by $C(X) = kX^a$, where k is a constant and $a (< 1)$ is the economies of scale parameter. Hence, $C_n = kX_n^a$ is the cost of expansion of size X_n , and, for the n th expansion,

$$V_{t_n}(K_1) = kC + E_{t_n} \left[e^{-r(T_{n+1}-t_n)} V_{T_{n+1}}(K_1) \right] \quad (4)$$

Now at time T_l , the total costs (TC) incurred are the actual cost of expansion (from initial capacity position of K_0 to the new capacity position of K_l), because of start of the expansion project; and the total cost of all the future expansion discounted at time T_l .

$$T \in E^{-r} \left[\sum_{i=1}^T V_i(K) \right] \quad (5)$$

where $X_1 = K_1 - K_0$ is the size of the first capacity expansion; and cost of continuing from the second capacity expansion is first discounted to time T_1 and then the total cost at time T_1 including the cost of expansion is discounted to time zero. In Equation (5), we note that the expected discount factor can be evaluated independently of $V_i(K)$ is possible because of the underlying independent increments in the demand model. Now, if we keep expanding the $V_i(K)$ term in Equation (5) using Equation (4), then the expression for the expansion cost can be written as a telescoping infinite series of costs:

$$T = E \left[e^{-rT} \left(\sum_{i=1}^T X_i E^{-rT_i} \right) \right] = \left\{ k + \frac{a}{2} k + \frac{a}{2} k + \dots \right\} \left\{ e^{-rT} + e^{-rT_1} + \dots \right\}$$

Now it can be shown that the total cost equation is equivalent to:

$$f(p, v) = C \frac{k(1 - v^a) p^{-p}}{1 - v^{a-p}} + p \sqrt{\frac{\mu^2}{\sigma^4} - \frac{2r\mu}{\sigma^2}} \quad (6)$$

Hence, the optimization problem is to find the minimum infinite horizon cost of expansion given by Equation (6), under the constraint that the capacity shortages (from Equation (3)) in the expansion cycle cannot exceed a certain specified limit. The decision variables are the timing and the size factors of expansion, as explained earlier. The problem is formulated as:

$$\begin{aligned} \min_{p, v} \quad & f(p, v) = \frac{k(1 - v^a) p^{-p}}{1 - v^{a-p}} \\ \text{subject to:} \quad & g(p, v) \leq 1 \end{aligned} \quad (7)$$

4. Solution Methodology

The non-linear program (7) is inherently difficult because of the complex constraint expression. Since the constraint inequality involves integration of bivariate normal density functions, it is very difficult to apply the commonly used gradient-based solution methods. Hence, a derivative-free cutting plane algorithm (Bazaraa et al. [9]) was used for the problem. Important steps of the cutting plane algorithm as it applies to our problem instance are described here:

Initialization step: Select an initial feasible point $x_0 = (p_0, v_0)$

For each iteration, solve the Master Problem, which is given as

$$\begin{aligned} \text{Maximize } & z \\ \text{s.t. } & t \leq f_j(p) + v_j(u) - p \text{ for } j = 1, \dots, k \\ & u \geq 0 \end{aligned}$$

Let (z_k, u_k) be the optimal solution. Now using the optimal value of the penalty variable u_k , solve the sub-problem:

$$\text{Minimize } f(p, v) + u_k g(p, v)$$

Let $x_k = (p_k, v_k)$ be the optimal solution for the sub problem. Let $\theta(u_k) = (p_k, v_k)$

If $z_k = \theta(u_k)$ then stop. Otherwise continue with the master problem with added constraint: $z \leq f(p_k, v_k) + u_k g(p_k, v_k)$

Figure 2. Cutting plane algorithm steps.

Zangwill [10] proved the convergence of this algorithm in a finite number of steps.

5. Numerical Results

The capacity expansion problem (CEP) in Equation (7) was solved using the cutting plane method with the parameter values as: drift of 8%, volatility of 20%, discount rate of 15%, economies of scale parameter of 0.9, and lead-time of 1 year, with the specified service level of 95%. The initial feasible point was taken to be (1.01, 1.01984).

Initial numerical runs of the algorithm indicated an unbounded solution. Hence to test convergence of the algorithm in Figure 2, we added an artificial constraint $p \leq 2$ to each sub-problem. The successive iterations and the convergence of the cutting plane algorithm are summarized in Table 1.

Table 1. Results of the cutting plane algorithm applied to CEP.

Iteration	Constraint Added	Master problem solution (z, u)	Sub-problem solution (p, v)	Sub-problem optimal value θ
1	$Z \leq 3.0156 - 0.001U$	(3.0156, 0)	(2, 1.15272)	1.04207
2	$Z \leq 1.04272 + 3.579U$	(2.88, 5.1455)	(2, 2.08893)	1.39851
3	$Z \leq 1.35628 + 0.0083U$	(1.762, 48.94)	(1, 1.29867)	0.0331
4	$Z \leq 2.7977 - 0.0564U$	(1.539, 22.32)	(2, 3.177)	1.0751
5	$Z \leq 1.7725 - 0.0367U$	(1.431, 9.307)	(2, 2.34)	1.40
6	$Z \leq 1.4543 - 0.0073U$	(1.4075, 6.45)	(2, 2.22)	1.403
7	$Z \leq 1.407 - 0.00072U$	(1.403, 5.974)	(2, 2.17)	1.4027
8	$Z \leq 1.387 + 0.00265U$	(1.4031, 6.16)	(2, 2.186)	1.4031

As seen from Table 1, the minimum cost for the CEP is achieved at decision variable values $(p, v) = (2, 2.186)$. Since our feasible region was $1 \leq p \leq 2$ we see that the optimal solution is reached at the artificially imposed boundary level of one of the decision variables-- indicating an *unbounded solution* to the CEP.

To explore whether the unboundedness was caused by the approximations used to transform Equation(1) to Equation (2), the same capacity expansion problem was solved using simulation to evaluate the constraint in Equation (1). That is, instead of using the analytical expression of the service level developed through the approximations and use of barrier option valuation formulas, the service level was directly computed by simulation of the GBM process using Matlab. The expression for the infinite time horizon expansion cost was the same as that was used in the analytical solution (Equation (6)). The graphical solution for the CEP is given below:

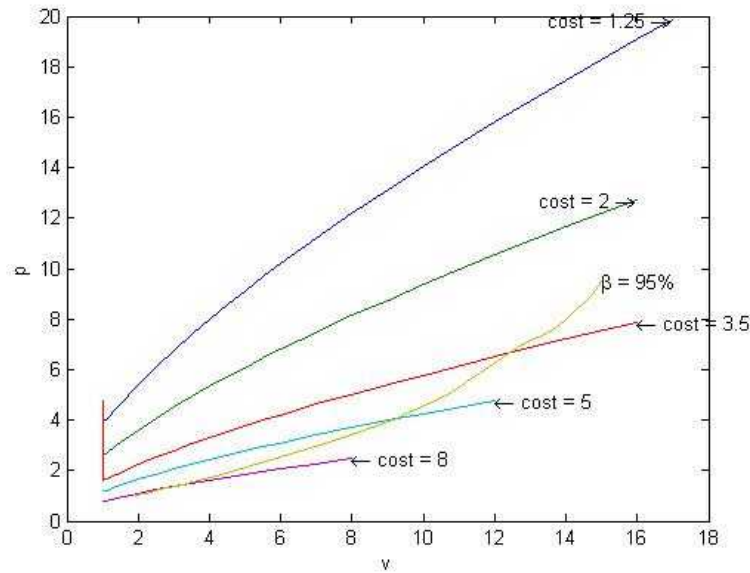


Figure 3. Graphical solution for the CEP via simulation for a 95% service level.

The feasible region is the area below the curve for the service level. The p - v cost contours are also plotted. As we see from Figure 3, we can go on decreasing the cost as we move away from the origin along the p - v curve for the service level. The simulation result was tested over wide range of parameter values, and every time we obtained similar plots. This leads us to believe that the problem under the current formulation is unbounded.

6. Conclusions and Future Work

From the formulation of the CEP and the numerical solution in section 5, we can conclude that if the service provider wants to start the capacity expansion when the demand has already crossed the current capacity position, he can do so by starting with any amount of initial shortage, provided that at each expansion the capacity increment is subsequently matched to satisfy the service level. In other words, if the service provider wants to start with a higher level of initial shortage, all that needs to be adjusted is the size of the capacity expansion so that the specified level of service is achieved (which can be accomplished by manipulating the constraint equation); moreover, this can be done without losing the minimum cost advantage (because, although the expansion cost increases in expansion size, the higher initial shortage will always pull it down). Simulation of the service level, instead of using analytical expectation, also leads to similar conclusions.

The possibility of achieving minimal cost with an acceptable service level by an unbounded initial shortage seems counter-intuitive. Hence, currently a reworking of the service level constraint is underway. Instead of considering the total shortages during the expansion cycle of random length, the distribution of instantaneous capacity shortage (in the same expansion cycle) is being considered.

7. References

1. Marathe, R.R. and Ryan, S.M., 2005a, "On the Validity of the Geometric Brownian Motion Processes," *The Engineering Economist*, 50(2), 159-192.
2. Manne, A.S., 1961, "Capacity Expansion and Probabilistic Growth," *Econometrica*, 29(4), 632-649.
3. Whitt, W., 1981, "The Stationary Distribution of a Stochastic Clearing Process," *Operations Research*, 29(2), 294-308.
4. Chaouch, A.B. and Buzacott, J.A., 1994, "The Effects of Lead Time on Plant Timing and Size," *Production and Operations Management*, 3(1), 38-54.
5. Bean, J.C., Higle, J.L. and Smith, R.L., 1992, "Capacity Expansion under Stochastic Demands," *Operations Research*, 40(2), S210-S216.
6. Ryan, S.M., 2004, "Capacity Expansion for Random Exponential Demand Growth with Lead Time," *Management Science*, 50(6), 740-748.
7. Marathe, R.R. and Ryan, S.M., 2005b, "Undercapacity as a Barrier Option: Evaluating the Service Level Constraint in Capacity Expansion," Under review with *Naval Research Logistics* (special issue on Applications of Financial Engineering).
8. Marathe, R.R. and Ryan, S.M., 2005c, "Capacity Expansion for Uncertain Demand with Initial Shortages," *Proceedings of the 14th IIE Research Conference*, May 14-18, 2004, Atlanta, Georgia.
9. Bazaraa, M.S., Sherali, H.D. and Shetty, C.M., 1993, "Nonlinear Programming: Theory and Algorithms," 2nd edition, John Wiley and Sons, New York.
10. Zangwill, W.I., 1969, "Nonlinear Programming: A Unified Approach," Prentice Hall, Englewood Cliffs, NJ.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.